



HAL
open science

Analyzing Clothing Layer Deformation Statistics of 3D Human Motions

Jinlong Yang, Jean-Sébastien Franco, Franck Hetroy-Wheeler, Stefanie Wuhrer

► **To cite this version:**

Jinlong Yang, Jean-Sébastien Franco, Franck Hetroy-Wheeler, Stefanie Wuhrer. Analyzing Clothing Layer Deformation Statistics of 3D Human Motions. ECCV 2018 - European Conference on Computer Vision, Sep 2018, Munich, Germany. pp.245-261, 10.1007/978-3-030-01234-2_15 . hal-01763706v2

HAL Id: hal-01763706

<https://inria.hal.science/hal-01763706v2>

Submitted on 25 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyzing Clothing Layer Deformation Statistics of 3D Human Motions

Jinlong Yang¹, Jean-Sébastien Franco¹, Franck Hétroy-Wheeler², and Stefanie Wuhrer¹

¹ Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK, 38000 Grenoble, France

* Institute of Engineering Univ. Grenoble Alpes

{jinlong.yang, jean-sebastien.franco, stefanie.wuhrer}@inria.fr

² ICube, University of Strasbourg, France

hetroywheeler@unistra.fr

Abstract. Recent capture technologies and methods allow not only to retrieve 3D model sequence of moving people in clothing, but also to separate and extract the underlying body geometry, motion component and the clothing as a geometric layer. So far this clothing layer has only been used as raw offsets for individual applications such as retargeting a different body capture sequence with the clothing layer of another sequence, with limited scope, *e.g.* using identical or similar motions. The structured, semantics and motion-correlated nature of the information contained in this layer has yet to be fully understood and exploited. To this purpose we propose a comprehensive analysis of the statistics of this layer with a simple two-component model, based on PCA subspace reduction of the layer information on one hand, and a generic parameter regression model using neural networks on the other hand, designed to regress from any semantic parameter whose variation is observed in a training set, to the layer parameterization space. We show that this model not only allows to reproduce previous retargeting works, but generalizes the data generation capabilities to other semantic parameters such as clothing variation and size, or physical material parameters with synthetically generated training sequence, paving the way for many kinds of capture data-driven creation and augmentation applications.

Keywords: Clothing Motion Analysis, 3D Garment Capture, Capture Data Augmentation and Retargeting

1 Introduction

Sequences showing the dense 3D geometry of dressed humans in motion are used in many 3D content creation applications. Two main approaches may be used to generate them, namely through physical simulation and by dense 3D motion capture. Physical simulators allow to generate realistic motions and cloth folding patterns based on given 3D models of the actor and the clothing, along with actor motion and cloth material parameters [7, 9, 17]. Dense 3D motion capture of human models has recently become possible at high spatial and temporal

resolution, e.g. using multi-camera systems [4, 10, 23]. While the captured data is unstructured, recent processing algorithms allow to track the captured geometry over time and to separate the actor’s body from the clothing [21, 38, 40].

While these works allow for the generation of accurate dense 3D motion sequences of human models with clothing, the content creation is expensive. Motion capture of the dense 3D geometry requires calibrated acquisition setups and the processing of the captured geometry requires heavy computation. Physical simulation requires artist generated models of actor and clothing and is typically computationally expensive.

In this work, we propose to leverage existing 3D motion sequences of dressed humans by performing a statistical analysis of the dynamically deforming clothing layer in order to allow for efficient synthesis of 3D motion. Performing statistical analysis on the clothing deformation is challenging for two main reasons. First, the motion of the clothing is influenced by numerous factors including the body shape and the motion of the underlying human as well as the cloth material. To allow for controlled synthesis of these effects, they must be explicitly modeled. Second, the geometry of the person with clothing may be significantly different from that of the underlying human body, e.g. in case of a dress, and the clothing may slide w.r.t. the human skin. Hence, computing assignments between the clothing layer and the body is complicated, especially across different subjects who wear the same type of clothing (e.g. shorts and T-shirt).

Two existing lines of work analyze the clothing layer to allow for synthesis. The first one analyzes multiple subjects wearing the same type of clothing based on simulated cloth deformations as training data [15, 37]. While these works allow for efficient synthesis, they can only be applied to simulated clothing, where the geometry is free of noise and explicit geometric correspondence information is available across all training data. Hence, the realism of the resulting cloth synthesis is limited by the quality of the cloth simulator used during training. The second line of work addresses this problem by analyzing the deformation of the clothing layer based on dense 3D motion capture data [21]. This work can handle noisy and unstructured data and the resulting model allows to change the body shape and motion under the clothing. However, the model can only be trained on one specific actor wearing a fixed outfit, and can hence not be used to synthesize changes in the clothing itself (e.g. changes of fit or materials).

This work combines the advantages of both methods by enabling to train from various motion data that can either be simulated or captured. To this end, we perform statistical analysis to model the geometric variability of the clothing layer. Our main contribution is that the proposed analysis is versatile in the sense that it can be used to train and regress on semantic parameters, thereby allowing to control e.g. clothing fit or material parameters of the synthesized sequences. Our statistical analysis models the deformation of the clothing layer w.r.t. the deformation behaviour of the underlying human actor represented using a statistical model of 3D body shape. For the analysis, we consider two fairly straightforward models. First, we model the layer variations with a linear subspace. Second, we model the variation of the clothing layer using a statistical

regression model, in order to capture some of the underlying causal dynamics in relatively simple form. Our experiments show the validity of the representation with qualitative and quantitative captured sequence reconstruction experiments based on these parameterizations. We further qualitatively demonstrate the value of our approach for three applications. First, following [21], we train from multiple sequences of the same actor in the same clothing acquired using dense 3D motion capture and use this to exchange body shape and motion. Second, following [15], we train from multiple simulated sequences showing the same actor in clothing of different materials and use this to change the material of the clothing. Third, to demonstrate the novelty of our approach, we train from multiple sequences of different actors in the same type of clothing acquired using dense 3D motion capture and use this to change the fit of the clothing.

2 Related work

This section reviews work on modeling the clothing layer. Furthermore, in case of captured 3D data, the models need to be processed to establish temporal coherence and to extract the clothing layer, and we provide a brief review of related literature.

Simulation-based modeling of clothing layer. To model the deformation of the clothing layer, a possible solution is direct physics-based simulation, for example with mass-spring systems [7, 9], continuum mechanics [34], or individual yarn structures [17]. The physical simulation models are complex and rely on numerous control parameters. Those parameters can be tuned manually, estimated from captures [32] or learned from perceptual experiments [30]. One line of works trains models on physics-based simulations using machine learning techniques, which subsequently allow for more efficient synthesis of novel 3D motion sequences of dressed humans [2, 15, 37]. In particular, these methods learn a regression from the clothing deformation to low-dimensional parameters representing body motion. These methods allow to modify the body shape, motion, and to alter the clothing. But the main disadvantage is that the methods are limited by the quality of the simulated synthetic training data. Since the simulation of complex clothing with multiple layers remains a challenging problem, this limitation restricts the model to relatively simple clothing. Our work addresses this problem by allowing to train from both simulated and captured sequences.

Capture-based modeling of clothing layer. Thanks to laser scanners, depth cameras, and multi-camera system, it is now possible to capture and reconstruct 3D human motion sequences as raw mesh sequences [4, 10, 23], and recent processing algorithms (reviewed in the following) allow to extract semantic information from the raw data. A recent line of work leverages this rich source of data by using captured sequences to learn the deformation of the clothing layer [22, 25]. Neophytou and Hilton [22] propose a method that trains from a single subject in fixed clothing and allows to change the body shape and motion after training. Pons-Moll *et al.* [25] extract the body shape and individual pieces of clothing from a raw capture sequence and use this information to trans-

fer the captured clothing to new body shapes. These methods allow to learn from complex deformations without requiring a physical model of the observations. The main disadvantage is that the model does not allow the modification of the clothing itself, such as the fit of the clothing and the cloth material. Our work addresses this problem by exploiting self-redundancies of the deformation to build a regression model from semantic sizing or material parameters to the clothing layer.

Processing raw dense 3D motion captures for human modeling. When captured 3D human motion sequences are used as input to our method, they first need to be processed to compute alignments and estimate the body under clothing. Recently, many solutions have been proposed for these challenging problems.

First approaches to process raw dense 3D motion captures of humans aimed to track human pose by fitting a 3D generic kinematic body-part model to captured 2D [28, 29, 31] or 3D data of the person [41]. To handle variation in body shape, some of these models adapt the size of rigid components of the skeleton or, taking this one step further, estimate shape parameters based on a human statistical shape space [5, 20, 21]. This accounts for the morphology of the captured human, for a closer data fit and a more accurate body estimate. Apart from such model fitting methods, other methods try to estimate the body shape by using convolutional neural networks [12, 11, 13]. Since all of these approaches assume tight clothing, they typically lead to inflation of the body estimate in the presence of wider clothing. To address this problem, recent methods propose to explicitly include wider clothing in the modeling. Instead of fitting a body shape as close as possible to the observation, the human body shape under clothing is captured by fitting the human model within the observed clothing contour from images [6, 26, 8] or the 3D clothing surfaces [16, 36, 38, 40]. These methods extract both the underlying body shape morphology and pose, as well as an explicit relative mesh representation of the clothing layer.

Aligning the clothing layer of a moving human is challenging because of the high deformation variability due to both the human pose and the non-rigidity of the cloth. To solve this problem some works combine reconstruction and tracking by deforming a detailed 3D model, typically obtained from a laser scan, to fit the capture data [33, 18, 14, 3]. These methods are usually applied to scenarios where captures are not dense enough to create a high-quality reconstruction, such as multi-camera systems with only few cameras. To prevent vertex drifting along the surface, some works exploit surface features to guide the template deformation [39]. Another line of works uses the estimated underlying body shape to align the clothed surface [27, 40, 25]. Our work leverages these recent processing methods for captured data to extract body shape estimations and alignments of the surface of the dressed person from given raw captured sequences.

3 Methodology

We propose a general framework to study the deformation of the clothing layer. First, we estimate the human body shape and extract the offset clothing layer

in a way that is robust to situations where the geometry of the dressed person differs significantly from the geometry of the human body, as in the case of a dress. A fuzzy vertex association from the clothing surface to the body surface is established, so that we can represent the clothing deformation as an offset mesh based on the body surface. Second, we use statistical analysis to analyze the geometric variability of the clothing layer to greatly reduce self-redundancies. Third, we show how to capture some of the underlying causal dynamics in relatively simple form, by modeling the variation of this clothing layer as a function of body motion as well as semantic variables using a statistical regression model.

The input to our method is a set of 3D sequences showing the dense geometry of a human in clothing performing a motion. These sequences may have been generated using physical simulation or motion capture set-ups. Before further processing, we require for each sequence an estimate of the underlying body shape and motion and an alignment of the clothing layer. Note that the clothing layer may optionally include the geometry of the body itself, i.e. show the body with clothing. If the sequences were generated using physical simulation, this information is typically readily available. For captured data, any of the previously reviewed methods may be used to compute this information. In this work, we estimate the underlying body shape using a recent method that explicitly takes wide clothing into account [38] and compute alignments of the complete deforming surface (i.e. the human in clothing) using an embedded deformation model based on Li *et al.* [18] without refining the deformation graph.

In the following, we denote the aligned sequences of the clothing layer by C_1, \dots, C_n and the corresponding sequences of underlying body shape estimates by B_1, \dots, B_n . Furthermore, let $C_{i,k}$ and $B_{i,k}$ denote the k -th frames of C_i and B_i , respectively. Thanks to the alignment, $C_{i,k}$ has the same number of corresponding vertices as $C_{j,l}$. Similarly, $B_{i,k}$ has the same number of corresponding vertices as $B_{j,l}$. While sequences C_i and C_j (and similarly B_i and B_j) may consist of different numbers of frames, C_i and B_i contain corresponding clothing layer and body estimate and therefore consist of the same number of frames.

The body estimates in sequence B_i can be expressed using a generative statistical body model that decouples the influence of identity and posture variation [21, 20, 24]. This allows to represent B_i using one vector β_i for identity information and a vector $\theta_{i,k}$ per frame for pose information. These generative models allow for two important modifications. First, the body shape of the actor can be changed while keeping the same motion by modifying β_i . Second, the body motion can be changed by modifying $\theta_{i,k}$ for each frame.

In this work, we use S-SCAPE as generative model [24], which uses the A-pose as standard pose θ_0 . S-SCAPE combines a linear space learned using principal component analysis (PCA) to represent variations due to identity with a linear blend skinning (LBS) model to represent variations in pose. Consider the j -th vertex $v_{i,k,j}^B$ of frame $B_{i,k}$. This vertex is generated by transforming the j -th vertex μ_j^B of the mean body shape in standard pose θ_0 as $v_{i,k,j}^B = T_j(\theta_{i,k})T_j(\beta_i)\mu_j^B$, where $T_j(\theta_{i,k})$ and $T_j(\beta_i)$ are (homogeneous) transformation matrices applying the transformations modeled by LBS and learned

by PCA. We can hence use S-SCAPE to define an operation called *unposing* in the following. This operation changes the pose of $\mathbf{B}_{i,k}$ to the standard pose $\boldsymbol{\theta}_0$ while maintaining body shape by replacing vertex $\mathbf{v}_{i,k,j}^{\mathbf{B}}$ for all j by

$$\tilde{\mathbf{v}}_{i,k,j}^{\mathbf{B}} = (\mathbf{T}_j(\boldsymbol{\theta}_{i,k}))^{-1} \mathbf{v}_{i,k,j}^{\mathbf{B}}. \quad (1)$$

3.1 Offset clothing layer extraction

We model the clothing layer as an offset from the body. To this end, we need to find corresponding vertices on the body mesh for each clothing vertex. Because $\mathbf{C}_1, \dots, \mathbf{C}_n$ and $\mathbf{B}_1, \dots, \mathbf{B}_n$ are temporally coherent, respectively, we can establish this correspondence on a single pair of frames $(\mathbf{C}_{i,k}, \mathbf{B}_{i,k})$ and propagate this information to all sequences. In practice, a pair of frames with few concavities is preferred because it enhances the robustness of the sparse association when created using a ray shooting method (see next paragraph). However to prove the generality of our approach, in our experiments, the association is simply estimated on the first frame of the first sequence. Since the following description is limited to a single pair of frames $(\mathbf{C}_{1,1}, \mathbf{B}_{1,1})$, for simplicity, we will drop frame and sequence index in this subsection.

\mathbf{C} and \mathbf{B} usually consist of a different number of vertices and have possibly significantly different geometry. Hence, a bijective association is in general not achievable. As our final goal is to model the deformation of the clothing layer using the body layer, our main interest is to find one or more corresponding vertices on \mathbf{B} for each vertex on \mathbf{C} . We achieve this by computing a sparse correspondence that is subsequently propagated to each vertex on \mathbf{C} using a probabilistic geodesic diffusion method. Note that unlike Pons-Moll *et al.* [25], our method works for difficult geometries such as skirts without manual intervention.

Sparse association For each vertex $\mathbf{v}_j^{\mathbf{B}}$ on \mathbf{B} we shoot a ray along the surface normal outwards the body. If there is an intersection $\mathbf{p}_j^{\mathbf{C}}$ with \mathbf{C} and the distance between $\mathbf{v}_j^{\mathbf{B}}$ and $\mathbf{p}_j^{\mathbf{C}}$ is within a threshold of 15cm, we search for the vertex $\mathbf{v}_i^{\mathbf{C}}$ on \mathbf{C} closest to $\mathbf{p}_j^{\mathbf{C}}$. Such pairs $(\mathbf{v}_i^{\mathbf{C}}, \mathbf{v}_j^{\mathbf{B}})$ are considered to be associated. If multiple body vertices are associated with the same clothing vertex, we only keep one pair per clothing vertex to put the same weight to each sparsely associated $\mathbf{v}_i^{\mathbf{C}}$. The pairs $(\mathbf{v}_i^{\mathbf{C}}, \mathbf{v}_j^{\mathbf{B}})$ are defined as *sparse association*.

Fuzzy dense association We now propagate the sparse association to every clothing vertex. Intuitively, if a clothing vertex $\mathbf{v}_i^{\mathbf{C}}$ is associated to a body vertex $\mathbf{v}_j^{\mathbf{B}}$ then there is a high probability that the neighboring vertices of $\mathbf{v}_i^{\mathbf{C}}$ should be associated to the neighboring vertices of $\mathbf{v}_j^{\mathbf{B}}$. Based on this idea, for any pair $(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_l^{\mathbf{B}}) \in \mathbf{C} \times \mathbf{B}$ we initialize the association probability $P(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_l^{\mathbf{B}})$ to be 0. Then we loop on all the sparse association pairs $(\mathbf{v}_i^{\mathbf{C}}, \mathbf{v}_j^{\mathbf{B}})$ and update the

association probability of any vertex pair $(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_l^{\mathbf{B}})$ according to:

$$P(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_l^{\mathbf{B}}) = P(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_l^{\mathbf{B}}) + \exp\left(-\left(r(\mathbf{v}_k^{\mathbf{C}}, \mathbf{v}_i^{\mathbf{C}}) + r(\mathbf{v}_l^{\mathbf{B}}, \mathbf{v}_j^{\mathbf{B}})\right)/\sigma^2\right), \quad (2)$$

where $r(\cdot)$ computes the squared geodesic distance between two vertices. In our implementation we set σ to $1cm$. To simplify the computation we only consider vertices $\mathbf{v}_k^{\mathbf{C}}$ and $\mathbf{v}_l^{\mathbf{B}}$ that lie within $3cm$ geodesic distance from $\mathbf{v}_i^{\mathbf{C}}$ and $\mathbf{v}_j^{\mathbf{B}}$. For the dense association, for each vertex on \mathbf{C} we choose a constant number n_f of vertices on \mathbf{S}_B that have the highest association probability values as associated vertices. We normalize the association probability to form fuzzy association weights, and store the indices of the n_f associations in a list I . This step does not only compute body vertex matches for previously unassociated clothing vertices but can also correct wrong matches from the sparse association and make the association more meaningful in situations where \mathbf{C} and \mathbf{B} differ significantly. This is illustrated in the case of a skirt on the right of Figure 1 (see also Section 4.1 for a discussion).

Offset representation of clothing layer Since we have established correspondence between \mathbf{C} and \mathbf{B} , we can now get the offset clothing layer by subtracting \mathbf{B} from \mathbf{C} . However, this Euclidean offset depends on the human pose and the global rotation. To account for this, we first unpose both \mathbf{B} and \mathbf{C} . The body estimate \mathbf{B} is unposed using Equation 1, and the clothing layer \mathbf{C} is unposed with the help of the fuzzy dense association by replacing vertex $\mathbf{v}_j^{\mathbf{C}}$ for all j by

$$\tilde{\mathbf{v}}_j^{\mathbf{C}} = \left(\sum_{i=1}^{n_f} \omega_i \mathbf{T}_{I_j[i]}(\boldsymbol{\theta})\right)^{-1} \mathbf{v}_j^{\mathbf{C}}, \quad (3)$$

where ω_i are the fuzzy association weights and $I_j[i]$ denotes the i -th entry of the index list I associated with vertex $\mathbf{v}_j^{\mathbf{C}}$. The offset of each clothing vertex is then obtained as:

$$\mathbf{d}_{i,j} = \tilde{\mathbf{v}}_i^{\mathbf{C}} - \tilde{\mathbf{v}}_j^{\mathbf{B}}, \quad \mathbf{d}_{i,j} \in \mathbb{R}^3, \quad (4)$$

where $(\tilde{\mathbf{v}}_i^{\mathbf{C}}, \tilde{\mathbf{v}}_j^{\mathbf{B}})$ form a fuzzily associated pair. We stack all the $\mathbf{d}_{i,j}$ from one frame pair (\mathbf{C}, \mathbf{B}) to form a single vector denoted by $\mathbf{d} \in \mathbb{R}^{3 \times n_f \times n_v}$, where n_v is the number of vertices in \mathbf{C} .

3.2 Clothing layer deformation space reduction

The deformation of the offset clothing layer is now encoded in \mathbf{d} . To reduce the self-redundancies in \mathbf{d} , we perform PCA on \mathbf{d} from all frame pairs $(\mathbf{C}_{i,k}, \mathbf{B}_{i,k})$. This allows for the clothing deformation to be represented by PCA coefficients $\boldsymbol{\alpha}_k$. Note we do not assume \mathbf{d}_k to form a Gaussian distribution. The purpose of PCA is only to reduce the dimensionality of the space, not to sample from it.

We would like to learn a mapping from semantic parameters of interest, denoted by $\boldsymbol{\gamma}$, to the clothing layer deformation. After obtaining a low dimensional

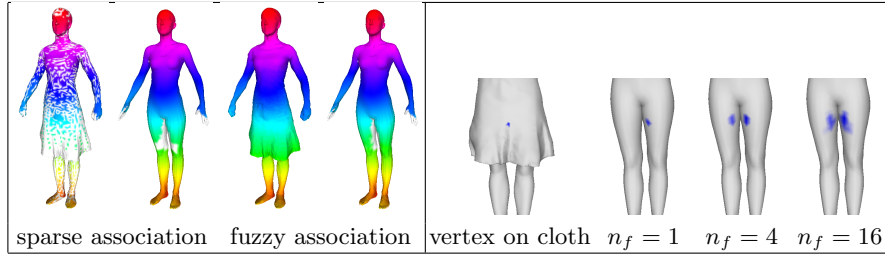


Fig. 1. Left: associations of clothing and body layer ($n_f = 1$), where color indicates the association. Right: a blue vertex on the skirt is associated to body vertices with different n_f . The intensity of the blue color is proportional to the association weight.

representation, this is equivalent to finding a mapping from γ to α . The PCA representation of the offsets successfully gets rid of self-redundancies in clothing layer. Furthermore, in PCA space, we can choose the number of principal components to use in order to balance the speed, storage, and quality.

3.3 Neural network for regression

To allow control of the offset clothing layer deformation, we study the relationship between its variation and semantic parameters γ , where γ can be body motion, clothing style, clothing material and so on. We treat this as a regression problem that learns the mapping from γ to α . Due to the nonlinearity of the problem itself and the potentially large sample size, we choose a fully connected two-hidden-layer neural network to train the regression, with the size of input layer equal to the dimensionality of the semantic parameters and the size of output layer equal to the number of principal component used. The sizes of the first and second hidden layers are 60 and 80, respectively. In our implementation, the neural network is implemented with OpenNN [1]. For each experiment, we set 20% of the frames from training data as validation frames. We choose mean square error as loss function, quasi-Newton method as optimization strategy, and stop the training once validation error starts to increase.

4 Method validation

To validate each step of our method, we train on small training sets consisting of a single sequence each ($n = 1$) using ten existing sequences of the Adobe [33] and Inria [38] datasets showing fast, large-scale motion in ample clothing as this is especially challenging to model. In all following experiments, body motion is parameterized by global speed, joint angles and joint angular speed. For offset clothing layer extraction, we show that we can extract the entire clothing layer regardless of the clothing geometry. Then we validate our PCA step to show that it greatly reduces the deformation space with acceptable reconstruction error. Finally, we validate the neural network regression by showing that both training error and testing error are satisfying.

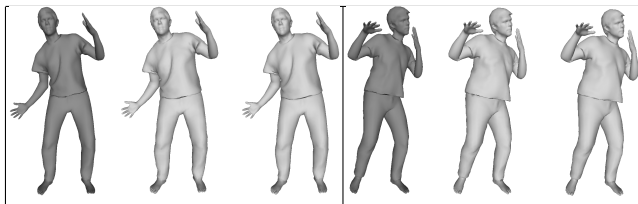


Fig. 2. Comparison to [25]. From left to right: original acquisition, transferred clothing layer with our method, and with [25]. Both methods produce very similar results.

4.1 Offset clothing layer extraction

We model the offset by first constructing a sparse correspondence between clothing and body, and then propagating the correspondence to each clothing vertex. Figure 1 (left) shows an example of the sparse and fuzzy association. Note that if we only use sparse association to store the information about clothing deformation, the information of the lower part of the dress is not recorded sufficiently.

The geometry of the clothing layer and the underlying human body differs significantly in the case of a skirt. Hence, $n_f = 1$ may not be meaningful and robust enough as having a single associated vertex is prone to form a seam in the middle of the front and the back faces of the skirt as some vertices around those areas are associated to the left thigh while neighboring ones are associated to the right thigh. Using higher n_f , such a skirt vertex is associated to both legs, therefore preventing seams. This is illustrated in Figure 1 (right).

We use our fuzzy association to directly transfer the offset clothing layer on data from Pons-Moll *et al.* [25]. Compared with their work, our method achieves similar results, shown in Figure 2, without the need for manual intervention.

4.2 PCA deformation space reduction

In our experiments, the dimension of the offset clothing layer vector generally varies from 20,000 to 80,000. To reduce this dimensionality, we perform PCA on the extracted offset of the clothing layer. To analyze how many PCs to keep, we reconstruct the sequence with different numbers of PCs. We compare the reconstruction against the original sequence by computing the average vertex position error. Table 1 gives errors per sequence for different numbers of principal components. Figure 3 visualizes the effect of increasing the number of PCs for one example. Such an analysis allows to choose the number of PCs to satisfy requirements on accuracy, speed or memory usage. In all following experiments, when training on a single subject with fixed clothing, we use 40 PCs, and when training on multiple subjects or multiple clothings, we use 100PCs as we found these datasets to contain more variation

4.3 Neural network regression

We validate our neural network by regressing 40 PCA coefficients to human body motion. Each sequence consists of 95-275 frames. We choose 20% of the frames

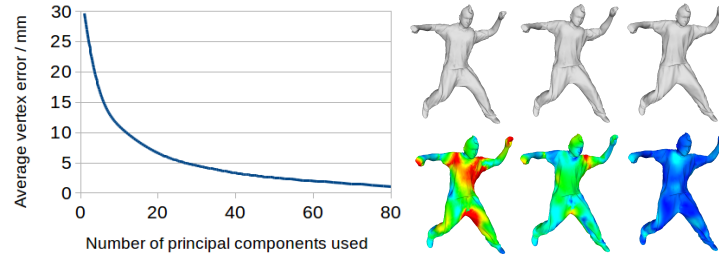


Fig. 3. Left: curve shows that the average reconstruction error drops when more principal components are used. Right: one example frame from *bouncing* sequence with the first row showing PCA reconstruction and the second row showing the error in color. From left to right 1 PC, 5 PCs and 40 PCs. Blue = 0mm, red ≥ 50 mm.

Seq	Bounc.	Crane	Mar. 1	Squat 1	Samba	Swing	s1 m. w.	s1 w. w.	s6 m. w.	s6 w. w.
1 PC	32.48	26.53	22.95	46.33	24.05	30.42	27.24	23.92	28.20	26.38
5 PC	22.52	14.87	13.94	18.94	14.75	19.68	13.73	13.28	12.94	12.25
40 PC	6.49	3.08	2.68	2.67	2.40	3.62	2.17	2.30	1.38	1.94
All PC	0.10	0.11	0.11	0.09	0.10	0.09	0.13	0.12	0.09	0.13

Table 1. Reconstruction error (mm) using different numbers of principal components.

from each sequence as testing data and the remaining 80% to be the training data. After training, we feed the motion parameters for all frames to the network and get 40 PCA coefficients for each frame to reconstruct the sequence. This reconstruction is then compared against the ground truth. Table 2 shows the quantitative error of the regression. The training error and the prediction error are generally low and close to the reconstruction error when using 40 principal components, which means our neural network regression is accurate and does not overfit the training data. Figure 4 shows the visual result of some examples of the regression error. Both training and prediction error are almost always low.

Seq	Bounc.	Crane	Mar. 1	Squat 1	Samba	Swing	s1 m. w.	s1 w. w.	s6 m. w.	s6 w. w.
E_{train}	9.95	4.71	3.54	6.57	4.27	7.72	10.87	9.09	7.71	5.97
E_{pred}	10.27	4.28	3.93	4.31	7.44	5.95	8.57	9.21	8.19	4.82

Table 2. Reconstruction error based on regression for each sequence.

5 Applications

This section shows the virtue of the proposed method by applying it to three scenarios. The first trains from multiple sequences of the same actor in the

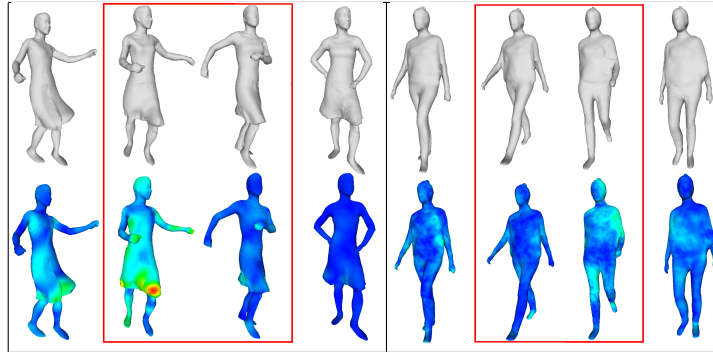


Fig. 4. Regression on two sequences. First row shows reconstruction. Second row shows vertex error on ground truth meshes. The two columns in the red box are predictions from testing frames; others are from training frames. Blue = $0mm$, red $\geq 50mm$.



Fig. 5. Change body shape. From left to right: original clothing mesh, estimated body, changed body, new clothing mesh.

same clothing and uses this to synthesize similar clothing on new body shapes and under new motions. The second trains from multiple simulated sequences showing the same actor in clothing of different materials and uses this to change the material of the clothing. The third application trains from multiple sequences of different actors in the same type of clothing and uses this to change the fit of the clothing. This entirely new way of synthesizing clothing is possible thanks to our regression to semantic parameters. For better visualizations of the results, refer to the supplemental material.

5.1 Clothing dynamics modeling

Change body shape After extracting the offset of the clothing layer, we can add this offset to any body shape under normalized pose and update the pose of the body with clothing using the relations of Equations 3 and 4. Figure 5 shows two examples of changing the body shape of a given motion sequence.

Change clothing dynamics In this part, we trained our regression model from multiple sequences of the same actor in the same clothing acquired using dense 3D motion capture. We use the regression model to learn the mapping from the body motion parameters to the PCA coefficients of the offset vectors. To synthesize new sequences, we feed new motion parameters to the model. Figure 6

shows examples of the resulting changes in the clothing dynamics. Note that realistic wrinkling effects are synthesized.

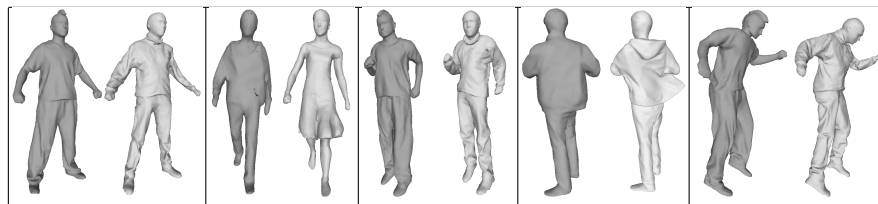


Fig. 6. Examples of changing clothing dynamics. The brighter gray meshes are not in the training data but generated by feeding the motion parameters of the darker gray meshes to the neural network trained on sequences containing the brighter gray clothes.

5.2 Clothing material modeling

This section shows how to model material parameters using our method. As material parameters are not readily available for captured data, we train from synthetic data generated using a state-of-art physical simulator [19]. For training, we simulate 8 sequences of the same garment pattern, worn by the same actor in a fixed motion, with varying materials. We choose a detailed garment pattern with garment-to-garment interaction during motion as this generates rich wrinkles that are challenging to model. The materials were generated using 39 parameters [35], and to allow for easier control of the parameters, we reduced their dimensionality to 4 using PCA before regressing from material parameter space to offset space. We used 7 materials to train our regression model, and left 1 material for testing. To avoid over-fitting to these 7 material points, we added a Gaussian random noise to the material parameters for all frames when training the regression. After training, we predicted the clothing layer from the motion parameters and new material parameters. Since for simulated data, segmented and aligned clothing and body meshes are available for each frame, our method uses this information. That is, we use the clothing layer directly and fit the S-SCAPE model to the mesh of the undressed body model used for simulation.

Figure 7 shows the comparison between our prediction and the ground truth for the test sequence. Note that a globally correct deformation is predicted even though the cloth deformation is far from the body shape. In spite of the globally correct deformation, our prediction lacks some detailed wrinkles. We suspect this detailed loss is due to dimension reduction on both material space and deformation space, as well as the limitation of the training data size. For qualitative validation, we randomly sampled material parameters in the PCA subspace and used them to synthesize new sequences. Figure 8 shows some examples. Note that visually plausible sequences are synthesized.

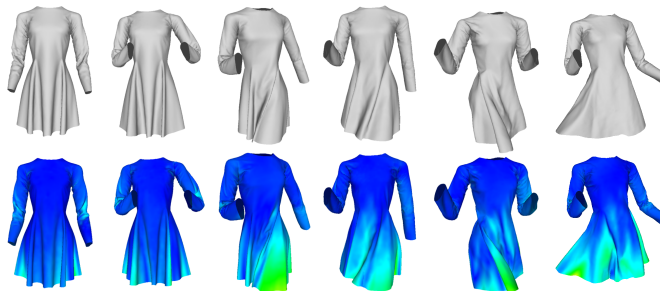


Fig. 7. Comparison to ground truth. First row: our predicted clothing deformation. Second row: ground truth colored with per-vertex error. Blue = 0cm, red = 10cm.



Fig. 8. Two synthesized sequences with new material parameters.

5.3 Clothing fit modeling

The proposed analysis is versatile in terms of the parameters of interest we wish to regress to. This allows for entirely new applications if sufficient training data is available. We demonstrate this by explicitly modeling the clothing size variation from acquisition data, which has not been done to the best of our knowledge. For training, we use 8 sequences of an extended version of the Inria dataset [38] of different subjects (4 male and 4 female) wearing different shorts and T-shirts while walking. These sequences are tracked with a common mesh topology. For each sequence, we manually assign a three dimensional vector to describe the size of the clothing, containing the width and length of the shorts and the size of the T-shirt. To model relative fit rather than absolute size, the sizes are expressed as ratio to corresponding measurements on the body. During training, to avoid over-fitting to these 8 sizing points, we add a Gaussian random noise to each size measurement. The regression learns a mapping from the body motion and size parameters to the PCA of the offsets. After training, new size parameters along with a motion allow to synthesize new sequences. Figure 9 shows modifications of the clothing fit on one frame of a sequence. Note that although our method learned certain clothing size variations, the three dimensions of our measurements are not completely separated, as e.g. the “large T-shirt” also introduces wider shorts. We believe this is caused by the limited size of training data. Since the regression also models body motion, our method

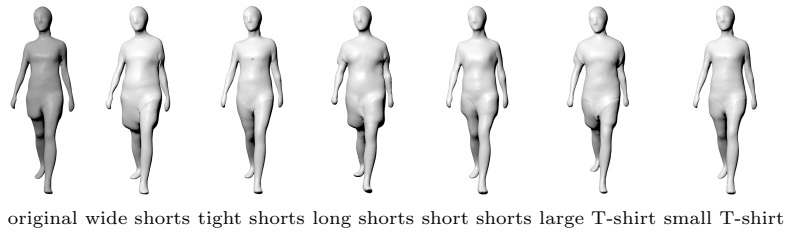


Fig. 9. Change the clothing fit shown on one frame of a sequence.

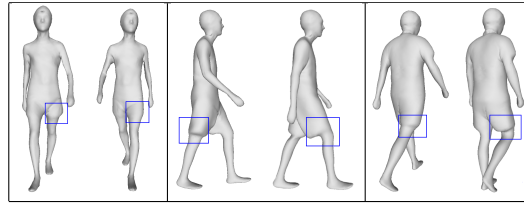


Fig. 10. Our approach captures dynamics caused by both clothing fit and body motion.

not only captures size variation, but also dynamic deformation caused by motion. Figure 10 shows examples of this.

6 Conclusion

In this paper we have presented a statistical analysis and modeling of the clothing layer from sets of dense 3D sequences of human motion. Our analysis shows PCA to be a suitable tool to compress the geometric variability information contained in the clothing layer. The regression component of our model is shown to properly capture the relation between layer variations and semantic parameters as well as the underlying motion of the captured body. This allows predictions of the clothing layer under previously unobserved motions, with previously unobserved clothing materials or clothing fits. Our model opens a large number of future possibilities. First, it can be extended to include more variability under different clothing worn by a large number of subjects. Second, more elaborate regression and clothing layer motion subspace models could be devised. Third, several semantic regression groups could be simultaneously considered.

7 Acknowledgement

Funded by France National Research grant ANR-14-CE24-0030 ACHMOV. We thank G. Pons-Moll for providing ClothCap [25] data, Kinovis platform³ for acquiring the extended Inria dataset, and F. Bertails-Descoubes and A. H. Rasheed for providing the cloth simulator and generating synthetic data.

³ <https://kinovis.inria.fr/>

References

1. Opennn library. <http://www.opennn.net/>
2. de Aguiar, E., Sigal, L., Treuille, A., Hodgins, J.K.: Stable spaces for real-time clothing. *ACM Transactions on Graphics* **29**(4) (2010)
3. de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.P., Thrun, S.: Performance capture from sparse multi-view video. *ACM Transactions on Graphics* **27**(3) (2008)
4. Allain, B., Franco, J.S., Boyer, E.: An efficient volumetric framework for shape tracking. In: *Conference on Computer Vision and Pattern Recognition* (2015)
5. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: SCAPE: shape completion and animation of people. *ACM Transactions on Graphics* **24**(3) (2005)
6. Bălan, A., Black, M.: The naked truth: Estimating body shape under clothing. *European Conference on Computer Vision* (2008)
7. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: *Conference on Computer Graphics and Interactive Techniques* (1998)
8. Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In: *European Conference on Computer Vision* (2016)
9. Bridson, R., Marino, S., Fedkiw, R.: Simulation of clothing with folds and wrinkles. In: *Symposium on Computer Animation* (2003)
10. Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., Hoppe, H., Kirk, A., Sullivan, S.: High-quality streamable free-viewpoint video. *ACM Transactions on Graphics* **34**(4) (2015)
11. Dibra, E., Jain, H., Öztireli, A.C., Ziegler, R., Gross, M.H.: Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In: *International Conference on 3D Vision* (2016)
12. Dibra, E., Jain, H., Öztireli, A.C., Ziegler, R., Gross, M.H.: Human shape from silhouettes using generative hks descriptors and cross-modal neural networks. In: *Conference on Computer Vision and Pattern Recognition* (2017)
13. Dibra, E., Öztireli, A.C., Ziegler, R., Gross, M.H.: Shape from selfies: Human body shape estimation using CCA regression forests. In: *European Conference of Computer Vision* (2016)
14. Gall, J., Stoll, C., De Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H.P.: Motion capture using joint skeleton tracking and surface estimation. In: *Conference on Computer Vision and Pattern Recognition* (2009)
15. Guan, P., Reiss, L., Hirshberg, D.A., Weiss, A., Black, M.J.: Drape: Dressing any person. *ACM Transactions on Graphics* **31**(4) (2012)
16. Hasler, N., Stoll, C., Rosenhahn, B., Thormählen, T., Seidel, H.P.: Estimating body shape of dressed humans. *Computers & Graphics* **33**(3) (2009)
17. Kaldor, J.M., James, D.L., Marschner, S.: Efficient yarn-based cloth with adaptive contact linearization. *ACM Transactions on Graphics* **29**(4) (2010)
18. Li, H., Adams, B., Guibas, L., Pauly, M.: Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics* **28**(5) (2009)
19. Li, J., Daviet, G., Narain, R., Bertails-Descoubes, F., Brown, G.E., Overby, M., Boissieux, L.: An implicit frictional contact solver for adaptive cloth simulation. *ACM Transactions on Graphics* **37**(4) (2018)
20. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.: SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics* **34**(6) (2015)

21. Neophytou, A., Hilton, A.: Shape and pose space deformation for subject specific animation. In: International Conference on 3D Vision (2013)
22. Neophytou, A., Hilton, A.: A layered model of human body and garment deformation. In: International Conference on 3D Vision (2014)
23. Newcombe, R.A., Fox, D., Seitz, S.M.: Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In: Conference on Computer Vision and Pattern Recognition (2015)
24. Pishchulin, L., Wuhler, S., Helten, T., Theobalt, C., Schiele, B.: Building statistical shape spaces for 3d human modeling. *Pattern Recognition* **67** (2017)
25. Pons-Moll, G., Pujades, S., Hu, S., Black, M.: Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics* **36**(4) (2017)
26. Rhodin, H., Robertini, N., Casas, D., Richardt, C., Seidel, H.P., Theobalt, C.: General automatic human shape and motion capture using volumetric contour cues. In: European Conference on Computer Vision (2016)
27. Shehu, A., Yang, J., Franco, J.S., Hétyroy-Wheeler, F., Wuhler, S.: Computing temporal alignments of human motion sequences in wide clothing using geodesic patches. In: International Conference on 3D Vision (2016)
28. Sidenbladh, H., Black, M.J., Fleet, D.J.: Stochastic tracking of 3d human figures using 2d image motion. In: European Conference on Computer Vision (2000)
29. Sigal, L., Balan, A.O., Black, M.J.: Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision* **87**(4) (2010)
30. Sigal, L., Mahler, M., Diaz, S., McIntosh, K., Carter, E., Richards, T., Hodgins, J.: A perceptual control space for garment simulation. *ACM Transactions on Graphics* **34**(4) (2015)
31. Sminchisescu, C., Triggs, B.: Estimating articulated human motion with covariance scaled sampling. *The International Journal of Robotics Research* **22**(6) (2003)
32. Stoll, C., Gall, J., De Aguiar, E., Thrun, S., Theobalt, C.: Video-based reconstruction of animatable human characters. *ACM Transactions on Graphics* **29**(6) (2010)
33. Vlasic, D., Baran, I., Matusik, W., Popović, J.: Articulated mesh animation from multi-view silhouettes. *ACM Transactions on Graphics* **27**(3) (2008)
34. Volino, P., Magnenat-Thalmann, N., Faure, F.: A simple approach to nonlinear tensile stiffness for accurate cloth simulation. *ACM Transactions on Graphics* **28**(4) (2009)
35. Wang, H., O'Brien, J.F., Ramamoorthi, R.: Data-driven elastic models for cloth: modeling and measurement. *ACM Transactions on Graphics* **30**(4) (2011)
36. Wuhler, S., Pishchulin, L., Brunton, A., Shu, C., Lang, J.: Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding* **127** (2014)
37. Xu, W., Umentani, N., Chao, Q., Mao, J., Jin, X., Tong, X.: Sensitivity-optimized rigging for example-based real-time clothing synthesis. *ACM Transactions on Graphics* **33**(4) (2014)
38. Yang, J., Franco, J.S., Hétyroy-Wheeler, F., Wuhler, S.: Estimation of human body shape in motion with wide clothing. In: European Conference on Computer Vision (2016)
39. Zaharescu, A., Boyer, E., Varanasi, K., Horaud, R.: Surface feature detection and description with applications to mesh matching. In: Conference on Computer Vision and Pattern Recognition (2009)

40. Zhang, C., Pujades, S., Black, M., Pons-Moll, G.: Detailed, accurate, human shape estimation from clothed 3d scan sequences. In: Conference on Computer Vision and Pattern Recognition (2017)
41. Zuffi, S., Black, M.J.: The stitched puppet: A graphical model of 3d human shape and pose. In: Conference on Computer Vision and Pattern Recognition (2015)